Asymptotic properties and variance estimators of the M-quantile regression coefficients estimators

BY ANNAMARIA BIANCHI

DMSIA, Università di Bergamo, annamaria.bianchi@unibg.it

NICOLA SALVATI

DSMAE, Università di Pisa, salvati@ec.unipi.it

SUMMARY

M-quantile regression is defined as a 'quantile-like' generalization of robust regression based on influence functions. The most appropriate assumption in the presence of gross errors in the data is that of independent not identically distributed (i.n.i.d.) regressors and errors. To the best of our knowledge, there is no existing theory for the asymptotic properties of the M-quantile regression coefficients estimators and the estimation of their variance in this context. The paper proves the consistency and asymptotic normality of the M-quantile regression coefficients estimators in both the i.i.d. and the i.n.i.d. settings. Estimators of the variance of the M-quantile regression coefficients appropriate to each setting are then proposed. Empirical results show that these estimators appear to perform well under different simulated scenarios.

Some key words: Influence function; M-estimation; Taylor expansion; Simulation experiments.

1. INTRODUCTION

Regression analysis is a standard tool for summarizing the average behaviour of a response variable y given a set of covariates x. It has been one of the most important statistical methods for applied research for many decades. However, the mean is not in general an adequate summary of a distribution, and so may be an inappropriate target of inference if one is, for example, interested in the extreme behaviour of y conditional on x. For this reason, Koenker & Bassett (1978) proposed the quantile regression (QR) model, which allows one to characterise the distribution of a response variable given a set of explanatory variables through models for the quantiles of this conditional distribution.

In the linear case, quantile regression leads to a family (or 'ensemble') of planes indexed by the value of the corresponding percentile coefficient. For each value of q in $(0, 1)$, the corresponding model shows how the q-th quantile of the conditional distribution of y given x, varies with x. In their seminal work, Koenker & Bassett (1978) formalized asymptotic properties of the least absolute deviation estimator of the QR model for independent observations.

Quantile regression can be viewed as a generalization of median regression. Similarly, expectile regression (Newey & Powell, 1987) is a 'quantile-like' generalization of mean, i.e. standard, regression. M-quantile (MQ) regression (Breckling & Chambers, 1988) extends this idea to a 'quantile-like' generalization of regression based on influence functions (M-regression). M-regression was introduced by Huber (1973) as a method of ensuring regression estimates that are robust against the presence of gross errors in the data. He also provided conditions that ensure the existence and asymptotic normality of the M-regression coefficients estimators for the case of fixed regressors and independent identically distributed (i.i.d.) errors. Yohai & Maronna (1979) provide similar results for stochastic i.i.d. regressors.

The aim of this paper is to study asymptotic properties of the M-quantile regression coefficients estimators. First we show how the results in Huber (1973) apply to this case (as already pointed out by Breckling & Chambers, 1988). Next we extend these results to the case of independent not identically distributed (i.n.i.d.) regressors and disturbances. This is an important case which to the best of our knowledge has not been discussed in the literature. This assumption is the most natural in the presence of gross errors in the data. Moreover it is the most appropriate one for relationships estimated using stratified cross-sectional data, e.g. when data are obtained from social and business Surveys, like the EU-SILC Surveys and the various national Labour Force Surveys. It is therefore very useful to have conditions which ensure that the familiar asymptotic properties of robust regression coefficients estimates hold under the i.n.i.d. case. We also propose covariance matrix estimators for both the i.i.d. and i.n.i.d. situations and prove their consistency.

The structure of the paper is as follows. In Section 2 we present the asymptotic properties of the M-quantile regression coefficients estimators and propose a covariance matrix estimator in the i.i.d. setting. The extension of these results to the i.n.i.d. case is provided in Section 3. These large sample approximations are then validated by a simulation study in Section 4. Finally, in Section 5 we summarize our main findings and provide directions for future research.

2. ASYMPTOTIC PROPERTIES AND COVARIANCE MATRIX ESTIMATION: THE I.I.D. CASE

Assume that the observations $\{(y_i, \mathbf{x}_i^T), i = 1, \ldots, n\}$ are generated by the linear model

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}_0 + u_i, \tag{1}$$

where $\{u_i\}$ is a sequence of independent identically distributed (i.i.d.) random variables with variance $\sigma^2 > 0$, $\boldsymbol{\beta}_0$ is a vector of unknown parameters and $\mathbf{x}_i$ are p-dimensional fixed regressors. We will denote $\mathbf{X}_n$ the $n \times p$ matrix with rows $\mathbf{x}_i^T$.

The M-quantile regression coefficient estimator (Breckling & Chambers, 1988) is defined as the vector $\hat{\boldsymbol{\beta}}_q$ which minimizes

$$\sum_{i=1}^{n} \rho_q \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right), \tag{2}$$

over $\boldsymbol{\beta}$, where ρ is the convex loss function associated with the M-quantile and ρ_q is defined as $\rho_q(u) := |q - I(u < 0)|\rho(u)$, for any $q \in (0, 1)$. The most common choices for ρ are $\rho(u) = |u|$, which corresponds to quantile regression (Koenker & Bassett, 1978), $\rho(u) = u^2$, which leads to expectile regression (Newey & Powell, 1987) and the Huber loss function $\rho(u) = (c|u| - \frac{1}{2}c^2)I(|u| > c) + \frac{1}{2}u^2 I(|u| \le c)$, which represents a compromise between the quantile and expectile loss functions. Since ρ is convex, the vector $\hat{\boldsymbol{\beta}}_q$ can equivalently be defined as the solution of the equations

$$\sum_{i=1}^{n} \psi_q \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \mathbf{x}_i = \mathbf{0}, \tag{3}$$

where $\psi_q(u) = d\rho_q(u)/du = |q - I(u < 0)|\psi(u)$, with $\psi(u) = d\rho(u)/du$. An iterative solution is needed here to obtain estimates of $\boldsymbol{\beta}_q$. The application of an iteratively reweighted least squares algorithm or the use of the Newton-Raphson algorithm then leads to a solution of (3). Theoretically the vector $\boldsymbol{\beta}_q$ is defined as the vector that minimizes the expectation of (2) or, equivalently, as the solution of the expected value of equation (3). In order to better understand the definition of $\boldsymbol{\beta}_q$, let $Q(\mathbf{x}_i, q) = \mathbf{x}_i^T\boldsymbol{\beta}_q$ denote the M-quantile of y_i. From location and scale equivariance of $Q(\mathbf{x}_i, q)$, it follows that $Q(\mathbf{x}_i, q) = \mathbf{x}_i^T\boldsymbol{\beta}_0 + Q(q)$, where $Q(q)$ is the q-th M-quantile of u_i. This implies that

$$\boldsymbol{\beta}_q = \boldsymbol{\beta}_0 + Q(q)\mathbf{e}_1, \quad \mathbf{e}_1 = (1, 0, \dots, 0)^T,$$

so that changing q only changes the intercept term in $\boldsymbol{\beta}_q$. On the other hand, when the scale of y_i also depends on $\mathbf{x}_i$, say $u_i = (\mathbf{x}_i^T\boldsymbol{\kappa})\varepsilon_i$, with ε_i independent of $\mathbf{x}_i$ and $\boldsymbol{\kappa} \in \mathbb{R}^p$, then

$$Q(\mathbf{x}_i, q) = \mathbf{x}_i^T\boldsymbol{\beta}_0 + Q(q)\mathbf{x}_i^T\boldsymbol{\kappa} = \mathbf{x}_i^T[\boldsymbol{\beta}_0 + Q(q)\boldsymbol{\kappa}],$$

where $Q(q)$ denotes the q-th M-quantile of ε_i. Hence in this case the slope coefficients in $\boldsymbol{\beta}_q$ also vary with q (since $\boldsymbol{\beta}_q = \boldsymbol{\beta}_0 + Q(q)\boldsymbol{\kappa}$). This case will be considered in more detail in Section 3.

The asymptotic theory for quantile regression and expectile regression models has already been studied by Koenker & Bassett (1978) and Newey & Powell (1987), respectively. The case of M-quantile regression with i.i.d. errors and fixed regressors can be easily derived from the results in Huber (1973), as pointed out in Breckling & Chambers (1988). For the sake of completeness we now state these results.

For $u_i(q) = (y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q)/\sigma$, $\psi'_{qi} = \psi'_q(u_i(q))$ and $\psi_{qi} = \psi_q(u_i(q))$, let

$$\mathbf{B}_n := \sigma^2 \frac{E[\psi_{q1}^2]}{(E[\psi'_{q1}])^2} \left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^T\right]^{-1}.$$

Consider the following assumptions:

ASSUMPTION 1. *The diagonal elements* γ_{ii} *of the projection matrix* $\mathbf{X}_n(\mathbf{X}_n^T\mathbf{X}_n)^{-1}\mathbf{X}_n^T$ *satisfy*

$$\max_{1\le i\le n} \gamma_{ii} \overset{n\to+\infty}{\longrightarrow} 0.$$

ASSUMPTION 2. *The function* ψ *is bounded and non-decreasing and possesses bounded derivatives up to the second order.*

ASSUMPTION 3. $E[\psi_{qi}] = 0$ *for all* i.

THEOREM 1. *Under Assumptions 1-3, for each* $q \in (0, 1)$ $\hat{\boldsymbol{\beta}}_q$ *exists in probability. Moreover*

$$\sqrt{n}\mathbf{B}_n^{-1/2}(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q) \overset{d}{\longrightarrow} N(\mathbf{0}, \mathbf{I}_p),$$

where $\mathbf{I}_p$ *is the identity matrix of size* p.

In order to use the previous theorem to build confidence intervals and hypothesis tests, a consistent estimator of the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_q$ is needed.

Assume that s is a consistent estimator for σ and define $\hat{u}_i(q) := (y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_q)/s$, $\hat{\psi}'_{qi} := \psi'_q(\hat{u}_i(q))$ and $\hat{\psi}_{qi} = \psi_q(\hat{u}_i(q))$. The analytical form of $\mathbf{B}_n$ suggests an estimator already proposed by Street et al. (1988) and now extended to covariance matrix of $\hat{\boldsymbol{\beta}}_q$:

$$\widehat{Var}_{SCR}(\hat{\boldsymbol{\beta}}_q) = \frac{1}{n}\hat{\mathbf{B}}_n = s^2 \frac{(n-p)^{-1}\sum_{i=1}^n \hat{\psi}_{qi}^2}{\left[n^{-1}\sum_{i=1}^n \hat{\psi}'_{qi}\right]^2} \Big[\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^T\Big]^{-1}. \quad (4)$$

An approximate confidence interval of level $(1-\alpha)$ for the hth coefficient at q is then

$$\hat{\beta}_{qh} \pm z_{(1-\alpha/2)}\sqrt{\widehat{Var}_{SCR}(\hat{\beta}_{qh})},$$

where $z_{(1-\alpha/2)}$ denotes the $(1-\alpha/2)$ quantile of the standard normal distribution.
In the next theorem we prove consistency of $\hat{\mathbf{B}}_n$.

THEOREM 2. *Under Assumptions 1-3,*

$$\hat{\mathbf{B}}_n - \mathbf{B}_n \xrightarrow{P} 0.$$

Proof. Since ψ'_{qi} is bounded and continuous in $\boldsymbol{\beta}$, the uniform law of large numbers yields

$$\sup_{\boldsymbol{\beta}} \left| \frac{1}{n}\sum_{i=1}^{n} \psi'_q\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) - E\left[\psi'_q\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right| \xrightarrow{P} 0.$$

Since $\hat{\boldsymbol{\beta}}_q \xrightarrow{P} \boldsymbol{\beta}_q$ and $s \xrightarrow{P} \sigma$, it follows from a slight modification of Lemma 2.6 in White (1980) that

$$\frac{1}{n}\sum_{i=1}^{n} \hat{\psi}'_{qi} \xrightarrow{P} E[\psi'_{q1}].$$

Analogously we can prove that

$$\frac{1}{n-p}\sum_{i=1}^{n} \hat{\psi}^2_{qi} \xrightarrow{P} E[\psi^2_{q1}].$$

So the result is a consequence of the continuous mapping theorem. □

The next result shows that using estimator (4) to test linear hypotheses and to construct confidence intervals is asymptotically correct.

Corollary. Under Assumptions 1-3,

$$\sqrt{n}\hat{\mathbf{B}}_n^{-1/2}(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_p)$$

and

$$n(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q)^T\hat{\mathbf{B}}_n(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q) \xrightarrow{d} \chi^2_p.$$

Proof. Consequence of Theorems 1 and 2 and Lemma 3.3 in White (1980). □

3. ASYMPTOTIC PROPERTIES AND COVARIANCE MATRIX ESTIMATION: THE I.N.I.D. CASE

In the previous Section we stated asymptotic properties of the M-quantile regression coefficients estimators in the case of i.i.d. errors and fixed regressors. However, these assumptions on the data are quite strong and do not cover many cases which arise in practice. For this reason we extend the results presented previously to the case of i.n.i.d. data, which is the most natural assumption in the presence of gross errors in the data. First we prove asymptotic normality and then we introduce a new covariance matrix estimator, which is a generalization of the previous one. In Section 4 we will compare the two estimators and see that this new estimator performs better in the presence of heteroskedasticity.

Assume that the observations $\{(y_i, \mathbf{x}_i^T), i = 1, \dots, n\}$ are generated by the linear model (1), with $\{(\mathbf{x}_i^T, u_i)\}$ a sequence of independent not (necessarily) identically distributed (i.n.i.d.) random vectors, and $\mathbf{x}_i$ a p-dimensional vector. The value u_i is a scalar error with variance σ_i^2 and $\boldsymbol{\beta}_0$ is a vector of unknown parameters. Notice that by assuming that $\{(\mathbf{x}_i^T, u_i)\}$ is an i.n.i.d. sequence, the case in which observations are obtained not from a controlled experiment but from a stratified cross-sectional sample is covered. Also covered is the case of fixed regressors with (possible) heteroskedastic errors.

Since the errors are no longer i.i.d., the definition of the M-quantile regression coefficient estimator needs to be modified from the definition given in the previous Section. Here $\hat{\boldsymbol{\beta}}_q$ is the vector that minimizes

$$\sum_{i=1}^{n} \rho_q \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma_i} \right), \tag{5}$$

over $\boldsymbol{\beta}$, or equivalently it is the solution of the equations

$$\sum_{i=1}^{n} \psi_q \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma_i} \right) \frac{\mathbf{x}_i}{\sigma_i} = \mathbf{0}. \tag{6}$$

For $u_i(q) = (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_q)/\sigma_i$, $\psi'_{qi} = \psi'_q(u_i(q))$ and $\psi_{qi} = \psi_q(u_i(q))$, let $\mathbf{W}_n^{-1}\mathbf{V}_n\mathbf{W}_n^{-1}$ be the asymptotic covariance of $\hat{\boldsymbol{\beta}}_q$ obtained by Taylor expansion (Binder, 1983) of (6) at $\hat{\boldsymbol{\beta}}_q = \boldsymbol{\beta}_q$ with

$$\mathbf{W}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} E[\psi'_{qi} \mathbf{x}_i \mathbf{x}_i^T]$$

$$\mathbf{V}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} E[\psi_{qi}^2 \mathbf{x}_i \mathbf{x}_i^T].$$

Notice that if the errors u_i are i.i.d. with variance σ^2 then we have that $y_i - \mathbf{x}_i^T \boldsymbol{\beta}_q = y_i - \mathbf{x}_i^T \boldsymbol{\beta}_0 - Q(q) = u_i - Q(q)$, where $Q(q)$ is the q-th M-quantile of u_i. It follows that $y_i - \mathbf{x}_i^T \boldsymbol{\beta}_q$ are i.i.d. random variables. If moreover the regressors are fixed, the matrices $\mathbf{W}_n$ and $\mathbf{V}_n$ simplify

$$\mathbf{W}_n = \sigma^{-2} E[\psi'_{q1}] \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T$$

$$\mathbf{V}_n = \sigma^{-2} E[\psi_{q1}^2] \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T,$$

and the asymptotic covariance matrix $\mathbf{B}_n$ follows directly.

The asymptotic theory will be developed under the following assumptions.

ASSUMPTION 4. *For each i there exists a finite constant M_{in} such that*

$$\frac{|\mathbf{x}_i|}{\sqrt{n}} \leq M_{in} \quad a.s. \quad \text{and} \quad \varepsilon_n := \max_{1 \leq i \leq n} M_{in}^2 \overset{n \to +\infty}{\longrightarrow} 0.$$

ASSUMPTION 5. *The matrices $\mathbf{V}_n$ and $\mathbf{W}_n$ are uniformly positive definite.*

ASSUMPTION 6. *There exists a finite positive constant Δ such that for all i $E|x_{ij}|^2 < \Delta$, $j = 1, \dots, p$.*

ASSUMPTION 7. *$E[\psi_{qi}\mathbf{x}_i] = 0$ for all i.*

ASSUMPTION 8. *The function ψ is bounded and non-decreasing and possesses bounded derivatives up to the second order.*

ASSUMPTION 9. *The variances $\sigma_i^2 = Var(u_i)$ are bounded away from zero, $\sigma_i^2 > \sigma_{MIN}^2$ for all i.*

Assumption 4 is a generalization of Assumption 1 in Section 2. It is used in the application of Lindeberg Central Limit Theorem. The uniform boundedness of the elements of $\mathbf{V}_n$ and $\mathbf{W}_n$ is guaranteed by Assumption 6. Assumption 6 together with Assumption 5 ensure that $\mathbf{V}_n$, $\mathbf{W}_n$ and their inverses are uniformly bounded and uniformly positive definite for n sufficiently large. The generality of the assumptions about $\mathbf{V}_n$ and $\mathbf{W}_n$ is necessary because they are not required to converge to any particular limit. This case is particularly important in the sampling situation, where the investigator usually cannot control the experiment to ensure the convergence of these matrices to some limit. Assumption 7 is an identifiability condition; it guarantees that the expectation of (5) reaches its minimum at the true value $\boldsymbol{\beta}_q$. When $E[\psi_{qi}] = 0$, independence of $u_i(q)$ and $\mathbf{x}_i$ is sufficient but not necessary for Assumption 7. Assumption 7 covers also the case in which $E[\psi_{qi}|\mathbf{x}_i] = 0$, thus allowing heteroskedasticity of the form $E[\psi_{qi}^2|\mathbf{x}_i] = g(\mathbf{x}_i)$, where g is a known function. Finally Assumption 8 is technically convenient. However we believe that the existence of higher order derivatives of ψ is not essential for the result to hold.
The next theorem demonstrates the asymptotic normality of $\hat{\boldsymbol{\beta}}_q$. Its proof is an extension of the proof of the asymptotic normality of $\boldsymbol{\beta}$ of M-regression in Huber (1973).

THEOREM 3. *Under Assumptions 4-9, for each $q \in (0, 1)$ $\hat{\boldsymbol{\beta}}_q$ exists in probability. Moreover*

$$\sqrt{n}\mathbf{V}_n^{-1/2}\mathbf{W}_n(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_p).$$

Proof. The matrix $\mathbf{W}_n$ is positive definite for n sufficiently large. Thus we can define the symmetric positive definite matrix $\mathbf{W}_n^{-1/2}$ such that $(\mathbf{W}_n^{-1/2})^2 = \mathbf{W}_n^{-1}$. The elements of $\mathbf{W}_n^{-1/2}$ are uniformly bounded under Assumptions 5, 6 and 9.
Without loss of generality we assume $\boldsymbol{\beta}_q = 0$ and consider the transformation

$$\mathbf{z}_i = \frac{1}{\sigma_i\sqrt{n}}\mathbf{W}_n^{-1/2}\mathbf{x}_i \qquad \boldsymbol{\beta}^* = \sqrt{n}\mathbf{W}_n^{1/2}\boldsymbol{\beta}.$$

Then $\hat{\boldsymbol{\beta}}_q^* = \sqrt{n}\mathbf{W}_n^{1/2}\hat{\boldsymbol{\beta}}_q$ is a solution of

$$\sum_{i=1}^{n} \psi_q\left(u_i(q) - \mathbf{z}_i^T\hat{\boldsymbol{\beta}}_q^*\right)\mathbf{z}_i = \mathbf{0}.$$

The idea is to compare the zeros of the two random functions:

$$\Phi(\boldsymbol{\beta}^*) = -\sum_{i=1}^{n} \psi_q\left(u_i(q) - \mathbf{z}_i^T\boldsymbol{\beta}^*\right)\mathbf{z}_i$$

$$\Psi(\boldsymbol{\beta}^*) = \boldsymbol{\beta}^* - \sum_{i=1}^{n} \psi_{qi}\mathbf{z}_i.$$

The zero $\hat{\boldsymbol{\beta}}_q^*$ of Φ is corresponds to $\sqrt{n}\mathbf{W}_n^{1/2}\hat{\boldsymbol{\beta}}_q$, while the zero of Ψ is

$$\tilde{\boldsymbol{\beta}}_q^* = \sum_{i=1}^n \psi_{qi}\mathbf{z}_i.$$

Let $\mathbf{a} \in \mathbb{R}^p$ such that $|\mathbf{a}| = \mathbf{a}^T\mathbf{a} = 1$. Thanks to Assumption 4, the Lindeberg Theorem applied to $\mathbf{a}^T\tilde{\boldsymbol{\beta}}_q^*$ implies that $\tilde{\boldsymbol{\beta}}_q^*$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $\sum_{i=1}^n E[\psi_{qi}^2\mathbf{z}_i\mathbf{z}_i^T] = \mathbf{W}_n^{-1/2}\mathbf{V}_n\mathbf{W}_n^{-1/2}$.
We will show that $|\hat{\boldsymbol{\beta}}_q^* - \tilde{\boldsymbol{\beta}}_q^*| = o_P(1)$, so that they follow the same law. Obviously, this result implies that $\sqrt{n}\hat{\boldsymbol{\beta}}_q = \mathbf{W}_n^{-1/2}\hat{\boldsymbol{\beta}}_q^*$ is asymptotically normal with mean zero and covariance matrix $\mathbf{W}_n^{-1}\mathbf{V}_n\mathbf{W}_n^{-1}$. Performing a Taylor expansion of $\mathbf{a}^T\Phi(\boldsymbol{\beta}^*)$, we have

$$\mathbf{a}^T\Phi(\boldsymbol{\beta}^*) = -\mathbf{a}^T\left[\sum_{i=1}^n \psi_{qi}\mathbf{z}_i + \sum_{i=1}^n \psi'_{qi}\mathbf{z}_i(\mathbf{z}_i^T\boldsymbol{\beta}^*) + \frac{1}{2}\sum_{i=1}^n \psi''_q(u_i(q) - \theta\mathbf{z}_i^T\boldsymbol{\beta}^*)\mathbf{z}_i(\mathbf{z}_i^T\boldsymbol{\beta}^*)^2\right]$$

with $0 < \theta < 1$. Since $\sum_i E[\psi'_{qi}\mathbf{z}_i\mathbf{z}_i^T] = \mathbf{I}_p$, we have that

$$\begin{aligned}
&\mathbf{a}^T[\Phi(\boldsymbol{\beta}^*) - \Psi(\boldsymbol{\beta}^*)] = \\
&= \mathbf{a}^T\sum_{i=1}^n[\psi'_{qi}\mathbf{z}_i\mathbf{z}_i^T - E(\psi'_{qi}\mathbf{z}_i\mathbf{z}_i^T)]\boldsymbol{\beta}^* - \frac{1}{2}\sum_{i=1}^n \psi''_q(u_i(q) - \theta\mathbf{z}_i^T\boldsymbol{\beta}^*)(\mathbf{z}_i^T\mathbf{a})(\mathbf{z}_i^T\boldsymbol{\beta}^*)^2 \\
&=: A + B.
\end{aligned}$$

In what follows we show that $\Phi - \Psi$ is uniformly bounded in a neighborhood of $\boldsymbol{\beta}^* = \mathbf{0}$, more precisely on sets of the form $\Gamma = \{(\boldsymbol{\beta}^*, \mathbf{a}) : |\boldsymbol{\beta}^*|^2 \leq K, |\mathbf{a}|^2 = 1\}$. Introduce the notation $r_i := \mathbf{z}_i^T\mathbf{a}$ and $t_i := \mathbf{z}_i^T\boldsymbol{\beta}^*$ and consider the first term

$$E(A^2) = \sum_{i=1}^n E(\mathbf{a}^T[\psi'_{qi}\mathbf{z}_i\mathbf{z}_i^T - E(\psi'_{qi}\mathbf{z}_i\mathbf{z}_i^T)]\boldsymbol{\beta}^*)^2 \leq \sum_{i=1}^n E(\psi'_{qi}r_i t_i)^2.$$

Notice that thanks to Assumptions 5 and 6, $|\mathbf{z}_i|^2 \leq \eta M_{in}^2/\sigma_{MIN}^2\ a.s.$, where η is a lower bound for the eigenvalues of $\mathbf{W}_n^{-1/2}$. Since $|\psi'_q(u)| \leq C'$ we have that

$$E(A^2) \leq \frac{C'\eta}{\sigma_{MIN}^2}|\boldsymbol{\beta}^*|^2\varepsilon_n\sum_{i=1}^n E[\psi'_{qi}|\mathbf{z}_i|^2] \leq C'\eta K\varepsilon_n p =: K_1\varepsilon_n.$$

Given $\delta > 0$, Markov's inequality implies that

$$P\left(|A| \geq \left(\frac{2}{\delta}K_1\varepsilon_n\right)^{1/2}\right) \leq \frac{\delta}{2},$$

for all $(\boldsymbol{\beta}^*, \mathbf{a}) \in \Gamma$. As far as the second term is concerned, assuming $|\psi_q''(u)| \leq C''$,

$$\begin{aligned} E|B| &\leq \frac{1}{2\sigma_{MIN}} C'' \eta \varepsilon_n^{1/2} E\left(\sum_i t_i^2\right) \\ &\leq \frac{1}{2\sigma_{MIN}} C'' \eta \varepsilon_n^{1/2} |\boldsymbol{\beta}^*|^2 tr\left(E\left[\sum_i \mathbf{z}_i \mathbf{z}_i^T\right]\right) \\ &\leq \frac{1}{2\sigma_{MIN}} C'' \eta K C_1 p \varepsilon_n^{1/2} =: K_2 \varepsilon_n^{1/2}, \end{aligned}$$

where we have used the fact that $tr\left(E[\sum_i \mathbf{z}_i\mathbf{z}_i^T]\right) \leq C_1 p$, since the matrix $E[\sum_i \mathbf{z}_i\mathbf{z}_i^T]$ is uniformly bounded by Assumption 6. Again, Markov's inequality yields that

$$P\left(|B| \geq \frac{2}{\delta} K_2 (\varepsilon_n)^{1/2}\right) \leq \frac{\delta}{2}.$$

It follows that

$$P(|\mathbf{a}^T[\Phi(\boldsymbol{\beta}^*) - \Psi(\boldsymbol{\beta}^*)]| < r) \geq 1 - \delta,$$

where $r := \left(\left(\frac{2}{\delta}K_1\right)^{1/2} + \frac{2}{\delta}K_2\right)(\varepsilon_n)^{1/2}$. Since this result holds simultaneously for all $\mathbf{a}$ with $|\mathbf{a}| = 1$,

$$P(|\Phi(\boldsymbol{\beta}^*) - \Psi(\boldsymbol{\beta}^*)| \leq r) \geq 1 - \delta. \tag{7}$$

Now, assuming $|\psi_q(u)| \leq C$, we see that

$$E|\tilde{\boldsymbol{\beta}}^*|^2 \leq C^2 C_1 p.$$

So from Markov's inequality it follows that $|\tilde{\boldsymbol{\beta}}^*|^2 \leq K/4$ with arbitrarily high probability, provided K is chosen large enough. Now notice that

$$|\boldsymbol{\beta}^* - \Phi(\boldsymbol{\beta}^*)| \leq |\tilde{\boldsymbol{\beta}}^*| + |\Psi(\boldsymbol{\beta}^*) - \Phi(\boldsymbol{\beta}^*)| \leq \frac{\sqrt{K}}{2} + r,$$

on the set $|\boldsymbol{\beta}^*| \leq \sqrt{K}$ with arbitrarily high probability. Since $\varepsilon_n \to 0$, for n sufficiently large r is smaller than $\sqrt{K}/2$, so that $|\boldsymbol{\beta}^* - \Phi(\boldsymbol{\beta}^*)| \leq \sqrt{K}$ for $|\boldsymbol{\beta}^*| \leq \sqrt{K}$. Applying Brouwer's fixed point theorem to the function $f(\boldsymbol{\beta}^*) := \boldsymbol{\beta}^* - \Phi(\boldsymbol{\beta}^*)$, we have that $\Phi(\boldsymbol{\beta}^*)$ has a zero $\hat{\boldsymbol{\beta}}^*$ such that $|\hat{\boldsymbol{\beta}}^*| \leq \sqrt{K}$.
Finally, substituting $\hat{\boldsymbol{\beta}}^*$ into (7), we obtain that $|\hat{\boldsymbol{\beta}}_q^* - \tilde{\boldsymbol{\beta}}_q^*| = o_P(1)$ and the result follows. □

To proceed further, we now need to define an estimator of the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_q$ for the i.n.i.d. situation. Assume that s_i is a consistent estimator for σ_i for each i and define $\hat{u}_i(q) := (y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_q)/s_i$, $\hat{\psi}'_{qi} := \psi'_q(\hat{u}_i(q))$ and $\hat{\psi}_{qi} = \psi_q(\hat{u}_i(q))$. The estimator that we propose is based on using the well known sandwich approach to estimate $\mathbf{W}_n$ and $\mathbf{V}_n$ directly from the sample data:

$$\widehat{Var}_{SAN}(\hat{\boldsymbol{\beta}}_q) = (n-p)^{-1} n \hat{\mathbf{W}}_n^{-1} \hat{\mathbf{V}}_n \hat{\mathbf{W}}_n^{-1} \tag{8}$$

where

$$\hat{\mathbf{W}}_n = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{s_i^2}\hat{\psi}'_{qi}\mathbf{x}_i\mathbf{x}_i^T,$$

$$\hat{\mathbf{V}}_n = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{s_i^2}\hat{\psi}^2_{qi}\mathbf{x}_i\mathbf{x}_i^T.$$

This estimator is based on White (1980) and takes into account the stochastic nature of the regressors, possible heteroskedasticity of the errors and possible non independence of errors and regressors. Note that the factor $(n-p)^{-1}n$ in equation (8) ensures agreement with expression (4) when $\mathbf{x} = \mathbf{1}$.

We now show that the covariance matrix estimator (8) is consistent. This requires the following additional assumption.

ASSUMPTION 10. *There exist finite positive constants δ, Δ such that for all i $E|x_{ij}x_{ik}|^{1+\delta} < \Delta$, for $j, k = 1, \ldots, p$.*

THEOREM 4. *Under Assumptions 4-10,*

$$\hat{\mathbf{W}}_n^{-1}\hat{\mathbf{V}}_n\hat{\mathbf{W}}_n^{-1} - \mathbf{W}_n^{-1}\mathbf{V}_n\mathbf{W}_n^{-1} \xrightarrow{P} 0.$$

Proof. Let us first prove the consistency of $\hat{\mathbf{W}}_n$. Its (j,k) component is given by $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{s_i^2}\hat{\psi}'_{qi}x_{ij}x_{ik}$. Notice that

$$|\frac{1}{\sigma_i^2}\psi'_{qi}x_{ij}x_{ik}| \leq \frac{C'}{\sigma^2_{MIN}}|x_{ij}x_{ik}| =: m(\mathbf{x}_i),$$

with $E|m(\mathbf{x}_i)|^{1+\delta} = (C'/\sigma^2_{MIN})^{1+\delta}E|x_{ij}x_{ik}|^{1+\delta} < +\infty$ by Assumption 10. From Theorem A.5 in Hoadley (1971) and thanks to the continuity of ψ'_q we have that

$$\sup_{\boldsymbol{\beta}}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\psi'_q\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma_i}\right)x_{ij}x_{ik} - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}E\left[\psi'_q\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma_i}\right)x_{ij}x_{ik}\right]\right| \xrightarrow{P} 0.$$

Since $\hat{\boldsymbol{\beta}}_q \xrightarrow{P} \boldsymbol{\beta}_q$ and $s_i \xrightarrow{P} \sigma_i$, it follows from a slight modification of Lemma 2.6 by White (1980) that $|\hat{\mathbf{W}}_n - \mathbf{W}_n| \xrightarrow{P} 0$. Similarly the convergence of $\hat{\mathbf{V}}_n$ to $\mathbf{V}_n$ can be proved. Since $\mathbf{W}_n$ is uniformly positive definite by Assumption 5, thanks to Proposition 2.30 in White (2001) we obtain the result. □

As the following Corollary demonstrates, the variance estimator (8) leads to asymptotically valid tests of linear hypotheses and confidence intervals.

Corollary. Under Assumptions 4-10,

$$\sqrt{n}\hat{\mathbf{V}}_n^{-1/2}\hat{\mathbf{W}}_n(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_p)$$

and

$$n(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q)^T\hat{\mathbf{W}}_n^{-1}\hat{\mathbf{V}}_n\hat{\mathbf{W}}_n^{-1}(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q) \xrightarrow{d} \chi^2_p.$$

Proof. Consequence of Theorems 3 and 4 and Lemma 3.3 by White (1980). □

Technically, the previous results do not hold for the Huber proposal 2 influence function because Assumption 8 is not satisfied. The next theorem shows that in fact they still hold.

THEOREM 5. *Under Assumptions 4-7, 9 and 10, with* $\psi(u) = uI(|u| \leq c) + c \cdot sgn(u)I(|u| > c)$

$$\hat{\mathbf{W}}_n^{-1}\hat{\mathbf{V}}_n\hat{\mathbf{W}}_n^{-1} - \mathbf{W}_n^{-1}\mathbf{V}_n\mathbf{W}_n^{-1} \xrightarrow{P} 0.$$

Proof. Convergence of $\hat{\mathbf{V}}_n$ to $\mathbf{V}_n$ follows as in Theorem 4. Since in the present case ψ' is not continuous we have to proceed differently to prove consistency of $\hat{\mathbf{W}}_n$.
Since $|s_i - \sigma_i| \xrightarrow{P} 0$ and $\sigma_i > \sigma_{MIN}$ for any i, $|1/s_i^2 - 1/\sigma_i^2| \xrightarrow{P} 0$, so that

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{s_i^2}\hat{\psi}'_{qi}\mathbf{x}_i\mathbf{x}_i^T - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\hat{\psi}'_{qi}\mathbf{x}_i\mathbf{x}_i^T\right| \geq \delta\right) \leq P\left(\frac{C'}{n}\sum_{i=1}^{n}\left|\frac{1}{s_i^2} - \frac{1}{\sigma_i^2}\right||\mathbf{x}_i|^2 \geq \delta\right)$$
$$\leq \frac{C'\varepsilon}{\delta n}\sum_{i=1}^{n}E|\mathbf{x}_i|^2 + o_P(1) \leq \frac{\varepsilon\Delta C'}{\delta} + o_P(1),$$

where we have used Markov's inequality, Assumption 10 and the fact that $|\hat{\psi}'_{qi}| \leq C'$. The previous quantity tends to zero for $\varepsilon \to 0$ and $n \to +\infty$. Now notice that $\hat{\psi}'_{qi}$ differs from ψ'_{qi} only if $|y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q| \leq |\mathbf{x}_i^T[\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q]|$ or if $-\sigma_i c < y_i - x_i^T\boldsymbol{\beta}_q < x_i^T[\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q] - s_i c$ or if $\mathbf{x}_i^T[\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q] - s_i c < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < -\sigma_i c$ or if $\sigma_i c < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < \mathbf{x}_i^T[\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q] + s_i c$ or if $\mathbf{x}_i^T[\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q] + s_i c < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < \sigma_i c$. Therefore,

$$|\hat{\psi}'_{qi} - \psi'_{qi}| \leq I(|y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q| \leq |\mathbf{x}_i^T[\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q]|)$$
$$+I(-\sigma_i c < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < \mathbf{x}_i^T[\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q] - s_i c) + I(\mathbf{x}_i^T[\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q] - s_i c < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < -\sigma_i c)$$
$$+I(\sigma_i c < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < \mathbf{x}_i^T[\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q] + s_i c) + I(\mathbf{x}_i^T[\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q] + s_i c < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < \sigma_i c).$$

It follows that for any fixed $\delta > 0$ and any $\varepsilon > 0$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\hat{\psi}'_{qi}\mathbf{x}_i\mathbf{x}_i^T - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\psi'_{qi}\mathbf{x}_i\mathbf{x}_i^T\right| \geq \delta\right) \leq P\left(\frac{1}{n\sigma_{MIN}^2}\sum_{i=1}^{n}|\mathbf{x}_i|^2|\hat{\psi}'_{qi} - \psi'_{qi}| \geq \delta\right)$$
$$\leq P(\frac{1}{n\sigma_{MIN}^2}\sum_{i=1}^{n}|\mathbf{x}_i|^2\{I(|y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q| \leq \varepsilon|\mathbf{x}_i|) + I(-\sigma_i c < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < \varepsilon|\mathbf{x}_i| - \sigma_i c + c\varepsilon)$$
$$+I(-\varepsilon|\mathbf{x}_i| - \sigma_i c - c\varepsilon < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < -\sigma_i c) + I(\sigma_i c < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < \varepsilon|\mathbf{x}_i| + \sigma_i c + c\varepsilon)$$
$$+I(-\varepsilon|\mathbf{x}_i| + \sigma_i c - c\varepsilon < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < \sigma_i c)\} \geq \delta) + o_P(1)$$
$$\leq \frac{1}{\delta\sigma_{MIN}^2 n}\sum_{i=1}^{n}E[|\mathbf{x}_i|^2\{I(|y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q| \leq \varepsilon|\mathbf{x}_i|) + I(-\sigma_i c < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < \varepsilon|\mathbf{x}_i| - \sigma_i c + c\varepsilon)$$
$$+I(-\varepsilon|\mathbf{x}_i| - \sigma_i c - c\varepsilon < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < -\sigma_i c) + I(\sigma_i c < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < \varepsilon|\mathbf{x}_i| + \sigma_i c + c\varepsilon)$$
$$+I(-\varepsilon|\mathbf{x}_i| + \sigma_i - c\varepsilon c < y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q < \sigma_i c)\}] + o_P(1),$$

where the last inequality follows from Markov's inequality. By the dominated convergence theorem, the consistency of $\hat{\boldsymbol{\beta}}_q$ and s_i and the fact that y_i is an absolutely continuous random variable, the previous quantity converges to zero for $\varepsilon \to 0$ and $n \to +\infty$.

Applying the triangle inequality, we have

$$|\hat{\mathbf{W}}_n - \mathbf{W}_n| \leq \left|\hat{\mathbf{W}}_n - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\hat{\psi}'_{qi}\mathbf{x}_i\mathbf{x}_i^T\right| + \left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\hat{\psi}'_{qi}\mathbf{x}_i\mathbf{x}_i^T - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\psi'_{qi}\mathbf{x}_i\mathbf{x}_i^T\right| + \left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\psi'_{qi}\mathbf{x}_i\mathbf{x}_i^T - \mathbf{W}_n\right|.$$

Consistency of $\hat{\mathbf{W}}_n$ then follows from the law of large numbers. □

4. SIMULATION STUDY

In this Section we provide results from a simulation study that was used to evaluate the large sample approximations outlined in Sections 2 and 3 and to investigate the performance of the variance estimators (4) and (8) of the M-quantile regression coefficients. Data are generated under the following model

$$y_i = 1 + 2x_i + u_i \quad i = 1, \ldots, 1000$$

with x drawn from a uniform distribution over the interval $[0, 1]$ for each replication. Five different settings are considered for the individual effects u_i:

(a) Gaussian errors with mean 0 and standard deviation 0.16;
(b) 10% contaminated Gaussian errors where 90% of errors are generated from a normal distribution with mean 0 and standard deviation 0.16 and the remaining 10% of errors are generated from a normal distribution with mean 0 and standard deviation 0.8;
(c) heteroschedastic errors where $u_i = \varepsilon_i(1 + \sqrt{2x})$ and ε_is are generated from a normal distribution with mean 0 and standard deviation 0.16;
(d) strong heteroschedastic errors where $u_i = \varepsilon_i \exp\{3x\}$ and ε_is are generated from a normal distribution with mean 0 and standard deviation 0.16;
(e) Chi-squared errors with 3 degrees of freedom.

The above scenarios will enable us to evaluate the large sample approximations and the performance of the variance estimators both when the assumption of the Gaussian errors holds and when this assumption is violated. Indeed, the first setting considers a situation of 'regularly' noisy data. The second one, on the contrary, defines a situation of more noisy data with the likely presence of outlying observations. In the last three settings the assumptions of Gaussian errors model are violated. A measure of the heteroskedasticity strength is given by $\lambda = max(\sigma_i^2)/min(\sigma_i^2)$. Under scenario (c) λ is equal to 3, corresponding to weak heteroskedasticity, whereas it assumes the value 20.06 under case (d), denoting strong heteroskedasticity. Each scenario is independently simulated $T = 5000$ times and samples of size $n = 100$ are selected from the simulated population, by simple random sampling. For the M-quantile model the Huber proposal 2 influence function is used with $c = 1.345$. This value gives reasonably high efficiency in the normal case - it produces 95% efficiency when the standardized errors are normal - and still offers protection against outliers (Huber, 1981). As estimate of scale the median absolute deviation (MAD) of the residuals is taken:

$$s = \frac{MED\{|(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_q) - MED|y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_q||; i = 1, \ldots, n\}}{0.6745}. \quad (9)$$

This estimator is convergent for σ when the error terms are $N(0,\sigma^2)$ (see, for example, Van der Vaart, 1998). Under settings (c) and (d) the heteroskedasticity is of the form $\sigma_i = \sigma v^{-1/2}(\mathbf{x}_i)$, where $v(\cdot)$ is a non-negative function. The MAD estimator of σ then becomes

$$s = \frac{MED\{|(\frac{y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_q}{v^{1/2}(\mathbf{x}_i)}) - MED|\frac{y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_q}{v^{1/2}(\mathbf{x}_i)}||; i = 1, \ldots, n\}}{0.6745},$$

and the scale σ_i is estimated by $s_i = sv^{-1/2}(\mathbf{x}_i)$.

For each type of distribution and for each estimator's component $\hat{\beta}_k$, $k = 0, 1$ at $q = (0.10, 0.25, 0.50, 0.75, 0.90)$, Tables 1 and 2 report

- the Monte-Carlo variance,

$$S^2(\hat{\beta}_k) = \frac{1}{T}\sum_{t=1}^{T}(\hat{\beta}_k^{(t)} - \bar{\beta}_k)^2,$$

  where $\hat{\beta}_k^{(t)}$ is the estimated M-quantile coefficient for the tth replication and $\bar{\beta}_k = T^{-1}\sum_{t=1}^{T}\hat{\beta}_k^{(t)}$;
- the estimated variance averaged over the simulations

$$\hat{S}^2(\hat{\beta}_k) = \frac{1}{T}\sum_{t=1}^{T}\widehat{Var}(\hat{\beta}_k);$$

- the coverage rate ($CR\%$) of nominal 95 per cent confidence intervals and its mean length. The coverage of these intervals is defined by the number of times the interval $\beta_k \pm 2\sqrt{\widehat{Var}(\hat{\beta}_k)}$ contains the 'true' population parameter.

Tables 1 and 2 summarize the results. In particular, Table 1 reports the results for the estimator (4), while Table 2 shows the behaviour of the sandwich-type variance estimator (8). Under the five scenarios the asymptotic variance estimators of the M-quantile regression coefficients provide a good approximation to the true variances. In particular, under scenarios (a), (b) and (e) the two estimators show similar performance in terms of point estimation and coverage rate, but estimator (8) has wider confidence intervals than the estimator (4). On the other hand, the sandwich-type variance estimator (8) has better performance than (4) under scenarios (c) and (d). As expected, the heteroskedasticity of the disturbances of the linear model used to generate the data in these cases causes (4) to become inconsistent.

We now turn to the examination of asymptotic normality of the M-quantile regression coefficients estimators. Figures 1 and 2 present normal probability plots of these estimators under the Gaussian and Chi-squared scenarios at quantiles $q = (0.25, 0.50, 0.75)$. The plots for other cases are not reported here, but are available from the authors upon request. It is observed that under the four scenarios a normal approximation of the distribution of the M-quantile regression coefficients estimators is reasonable. This is also confirmed by values W of the Shapiro-Wilk's test with a few exceptions. Under the setting based on Chi-squared errors there are some departures from normality for $q = 0.25$ and $q = 0.75$, but they are not severe and we believe that the normal approximation improves as soon as the sample size increases. This can also be supported by the fact that the coverage rates of normal confidence intervals for β_0 and β_1 are about 95% for all percentiles. Consequently, the construction of confidence intervals based on asymptotic normality seems to be correct. We therefore conclude that the proposed large sample approxima-

Errors		$\hat\beta_0$				$\hat\beta_1$			
		$S^2(\hat\beta_k)$	$\hat S^2(\hat\beta_k)$	$CR\%$	Mean length	$S^2(\hat\beta_k)$	$\hat S^2(\hat\beta_k)$	$CR\%$	Mean length
Gaussian	$q = 0.10$	0.0118	0.0116	94.0	0.412	0.0355	0.0347	94.0	0.715
	$q = 0.25$	0.0080	0.0078	95.5	0.344	0.0241	0.0235	95.3	0.596
	$q = 0.50$	0.0069	0.0068	95.0	0.321	0.0210	0.0204	95.3	0.557
	$q = 0.75$	0.0079	0.0078	95.0	0.343	0.0247	0.0235	95.0	0.596
	$q = 0.90$	0.0118	0.0115	94.0	0.412	0.0370	0.0347	93.7	0.714
Contamined	$q = 0.10$	0.0409	0.0356	93.1	0.693	0.1215	0.1072	92.8	1.203
	$q = 0.25$	0.0120	0.0120	95.7	0.423	0.0370	0.0359	95.3	0.734
	$q = 0.50$	0.0089	0.0090	95.9	0.369	0.0276	0.0271	95.9	0.641
	$q = 0.75$	0.0120	0.0119	95.3	0.422	0.0372	0.0357	95.2	0.732
	$q = 0.90$	0.0403	0.0358	93.3	0.694	0.1236	0.1077	93.1	1.205
Heteroskedastic	$q = 0.10$	0.0167	0.0116	90.0	0.412	0.0657	0.0348	85.3	0.716
	$q = 0.25$	0.0112	0.0078	91.0	0.344	0.0443	0.0235	86.6	0.557
	$q = 0.50$	0.0097	0.0068	91.0	0.321	0.0389	0.0204	85.6	0.557
	$q = 0.75$	0.0112	0.078	91.1	0.343	0.0458	0.0236	85.6	0.596
	$q = 0.90$	0.0168	0.0115	89.4	0.412	0.0693	0.0347	85.0	0.714
Strong Heteroskedastic	$q = 0.10$	0.0221	0.0115	84.2	0.412	0.1583	0.0346	64.6	0.715
	$q = 0.25$	0.0153	0.0078	85.5	0.344	0.1100	0.0235	65.5	0.597
	$q = 0.50$	0.0134	0.0068	85.2	0.321	0.0976	0.0204	66.1	0.557
	$q = 0.75$	0.0154	0.0078	85.9	0.343	0.1133	0.0235	65.0	0.596
	$q = 0.90$	0.0226	0.0115	85.0	0.411	0.1665	0.0346	63.7	0.714
Chi-squared	$q = 0.10$	0.0660	0.0611	92.8	0.959	0.1901	0.1835	95.0	1.663
	$q = 0.25$	0.1013	0.0948	92.4	1.198	0.2883	0.2850	95.5	2.078
	$q = 0.50$	0.1813	0.1659	91.7	1.585	0.5121	0.4985	95.5	2.749
	$q = 0.75$	0.4270	0.4358	94.3	2.548	1.3031	1.3100	95.0	4.420
	$q = 0.90$	1.2195	1.2600	93.3	4.213	3.7547	3.779	93.1	7.305

Table 1. *Empirical variances ($S^2(\hat\beta_k)$), estimated variances (4) averaged over the simulations ($\hat S^2(\hat\beta_k)$), coverage rate ($CR\%$) and average lengths of 95% confidence intervals on β_k. Sample size n=100.*

tions are suitable for approximate inference based on the estimators of the $\boldsymbol{\beta}_q$ coefficients of the M-quantile model.

Errors		$\hat{\beta}_0$				$\hat{\beta}_1$			
		$S^2(\hat{\beta}_k)$	$\hat{S}^2(\hat{\beta}_k)$	$CR\%$	Mean length	$S^2(\hat{\beta}_k)$	$\hat{S}^2(\hat{\beta}_k)$	$CR\%$	Mean length
	$q = 0.10$	0.0118	0.0123	93.0	0.415	0.0355	0.0368	93.1	0.725
	$q = 0.25$	0.0080	0.0080	95.0	0.344	0.0241	0.0239	94.9	0.598
Gaussian	$q = 0.50$	0.0069	0.0069	95.1	0.321	0.0210	0.0206	95.0	0.557
	$q = 0.75$	0.0079	0.0080	94.7	0.343	0.0247	0.0239	94.4	0.597
	$q = 0.90$	0.0118	0.0124	92.8	0.416	0.0370	0.0370	93.1	0.725
	$q = 0.10$	0.0409	0.0546	93.1	0.768	0.1215	0.1684	95.1	1.362
	$q = 0.25$	0.0120	0.0126	95.6	0.429	0.0370	0.0378	96.0	0.747
Contamined	$q = 0.50$	0.0089	0.0092	95.8	0.371	0.0276	0.0277	95.6	0.645
	$q = 0.75$	0.0120	0.0125	95.0	0.426	0.0372	0.0377	95.2	0.745
	$q = 0.90$	0.0403	0.0540	92.4	0.764	0.1236	0.1598	95.9	1.363
	$q = 0.10$	0.0167	0.0175	92.1	0.488	0.0657	0.0693	93.3	0.993
	$q = 0.25$	0.0112	0.0112	94.4	0.405	0.0443	0.0445	94.6	0.816
Heteroskedastic	$q = 0.50$	0.0097	0.0096	95.0	0.379	0.0389	0.0383	95.3	0.760
	$q = 0.75$	0.0112	0.0111	94.3	0.404	0.0458	0.0444	94.6	0.815
	$q = 0.90$	0.0168	0.0174	91.5	0.487	0.0693	0.0692	92.7	0.990
	$q = 0.10$	0.0221	0.0239	90.3	0.557	0.1583	0.1704	93.8	1.561
Strong	$q = 0.25$	0.0153	0.0153	93.5	0.470	0.1100	0.1123	95.1	1.297
Heteroskedastic	$q = 0.50$	0.0134	0.0132	94.9	0.442	0.0976	0.0969	95.0	1.211
	$q = 0.75$	0.0154	0.0153	94.0	0.471	0.1133	0.1121	95.1	1.296
	$q = 0.90$	0.0226	0.0233	90.6	0.552	0.1665	0.1684	93.7	1.554
	$q = 0.10$	0.0660	0.0627	91.9	0.964	0.1901	0.1883	95.0	1.676
	$q = 0.25$	0.1013	0.0961	91.2	1.200	0.2883	0.2890	95.8	2.087
Chi-squared	$q = 0.50$	0.1813	0.1688	90.5	1.591	0.5121	0.5089	95.8	2.768
	$q = 0.75$	0.4270	0.4646	92.6	2.587	1.3031	1.3960	94.9	4.519
	$q = 0.90$	1.2195	1.0112	91.5	4.488	3.7547	3.224	92.5	7.867

Table 2. *Empirical variances ($S^2(\hat{\beta}_k)$), estimated variances (8) averaged over the simulations ($\hat{S}^2(\hat{\beta}_k)$), coverage rate ($CR\%$) and average lengths of 95% confidence intervals on β_k. Sample size n=100.*

5. Discussion

M-quantile regression allows one to investigate the behaviour of the conditional M-quantile regression functions of a response variable in terms of a set of covariates. Like ordinary quantiles, the M-quantiles characterize a distribution, and so the M-quantile regression functions lead to a more complete picture of this relationship. However, little or no work has been done on the asymptotic properties of the M-quantile regression coefficients estimators and on the estimation

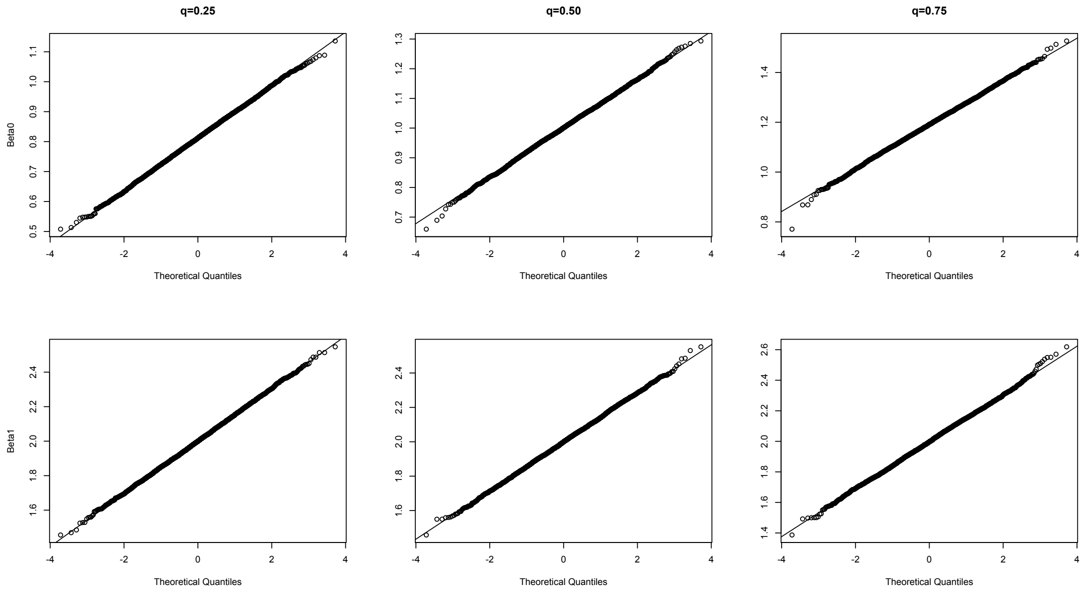

Fig. 1. Q–Q plots of estimates of $\boldsymbol{\beta}_q$ for $q = 0.25, 0.50, 0.75$ in the case of Gaussian errors.

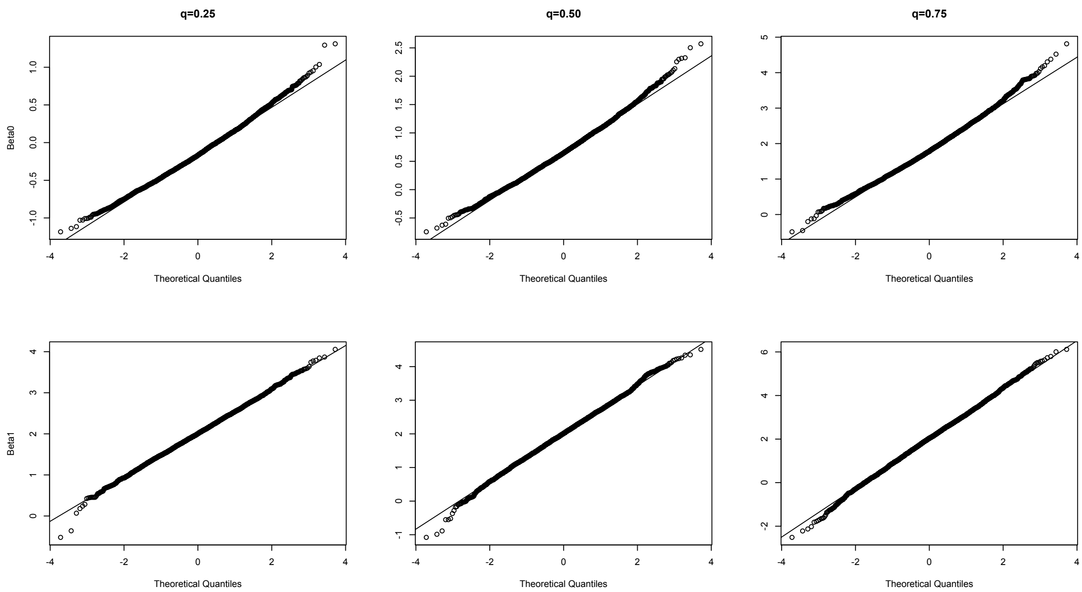

Fig. 2. Q–Q plots of estimates of $\boldsymbol{\beta}_q$ for $q = 0.25, 0.50, 0.75$ in the case of Chi-squared errors.

of their variances. This is particularly true for the i.n.i.d. case, which is appropriate when there are gross errors in the data.

In this paper we prove the consistency and asymptotic normality of the M-quantile regression coefficients estimators for this case. Moreover we propose two estimators of the asymptotic variance of the M-quantile regression coefficients estimator and we prove their consistency. The empirical results described in the previous Section provide empirical support for the asymptotic theory developed in Sections 2 and 3 and demonstrate that the variance estimators (4) and (8) are reasonably efficient over a wide range of error distributions.

The variance of the M-quantile regression coefficients could have also been estimated with a resampling method. In particular, it can be shown that a 'one-step' jackknife estimator of the variance of the M-quantile regression coefficients is asymptotically equivalent to the sandwich estimator (8). We have evaluated the performance of this 'one-step' jackknife estimator and have observed that it works as well as the sandwich-type variance estimator (8), except for some extreme quantiles where the 'one-step' jackknife estimator showed some under-coverage. For reasons of space, these results are not reported in the paper. However, they are available from the authors upon request.

By construction, M-quantile-based estimators are robust against outliers in the distributions of the random errors. However, the method does not take into account the presence of outliers in the auxiliary variables. In such cases, the M-quantile estimating equations (6) for $\boldsymbol{\beta}_q$ can be modified as

$$\sum_{i=1}^{n} \mathbf{d}_i \psi_q \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma_i} \right) \frac{\mathbf{x}_i}{\sigma_i} = \mathbf{0}, \tag{10}$$

where $\mathbf{d}_i$ is a diagonal matrix of weight functions $d(\mathbf{x}_i)$ which allows one to downweight the outliers in the auxiliary variables when estimating the parameters of the M-quantile model. In this context, we note that a function of the Mahalanobis distance can be used as the weight function $d(\mathbf{x}_i)$ when the x's are continuous (Sinha, 2004). If the auxiliary variables are discrete, Mallows weights can be used (see de Jongh et al., 1988).

Recently Pratesi et al. (2009) have extended the M-quantile regression model to nonparametric penalized spline regression, in the sense that the M-quantile regression functions do not have to follow any particular functional form, but can be left undefined and estimated from the data. It could be interesting to explore whether the variance estimators proposed in Sections 2 and 3 can also be used in this case. Here we note that the asymptotic results would then have to take into account the penalty term and the number and position of knots use in the spline approximation.

REFERENCES

BINDER, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review* **51**, 279–292.

BRECKLING, J. & CHAMBERS, R. (1988). M-quantiles. *Biometrika* **75**, 761–771.

DE JONGH, P., DE WET, T. & WELSH, A. (1988). Mallows-type bounded-influence-regression trimmed means. *Journal of the American Statistical Association* **83**, 805–810.

HUBER, P. (1973). Robust regression: Asymptotics, conjectures and monte carlo. *The Annals of Statistics* **1**, 799–821.

HUBER, P. J. (1981). *Robust Statistics*. John Wiley & Sons.

KOENKER, R. & BASSETT, G. (1978). Regression quantiles. *Econometrica* **46**, 33–50.

NEWEY, W. & POWELL, J. (1987). Asymmetric least squares estimation and testing. *Econometrica* **55**, 819–847.

PRATESI, M., RANALLI, M. G. & SALVATI, N. (2009). Nonparametric M-quantile regression via penalized splines. *Journal of Nonparametric Statistics* **21**, 287–304.

SINHA, S. (2004). Robust analysis of generalized linear mixed models. *Journal of the American Statistical Association* **99**, 451–460.

STREET, J., CARROLL, R. & RUPPERT, D. (1988). A note on computing robust regression estimates via iteratively reweighted least squares. *The American Statistician* **42**, 152–154.

VAN DER VAART, A. (1998). *Asymptotic Statistics*. Cambridge University Press.

WHITE, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and direct test for heteroskedasticity. *Econometrica* **48**, 817–838.

WHITE, H. (2001). *Asymptotic Theory for Econometricians*. Academic Press, San Diego.

YOHAI, V. J. & MARONNA, R. A. (1979). Asymptotic behavior of M-estimators for the linear model. *The Annals of Statistics* **7**, 258–268.